Computation of charge distribution and electrostatic potential in silicates with the use of chemical potential equalization models

Toon Verstraelen,[*,†] Sergey V. Sukhomlinov,[‡] Veronique Van Speybroeck,[†] Michel Waroquier,[†] and Konstantin S. Smirnov[*,‡]

Center for Molecular Modeling (CMM), Ghent University, Technologiepark 903, 9052 Ghent, Belgium, (Member of the QCMM Ghent-Brussels Alliance Group), and Laboratoire de Spectrochimie Infrarouge et Raman (LASIR), CNRS, Université Lille 1 – Sciences et Technologie, 59655 Villeneuve d'Ascq, France

E-mail: Toon.Verstraelen@ugent.be; Konstantin.Smirnov@univ-lille1.fr

Fax: +33 3 2043 6755

[*]To whom correspondence should be addressed
[†]Ghent University
[‡]Université Lille 1 – Sciences et Technologie

Abstract

New parameters for the electronegativity equalization model (EEM) and the split-charge equilibration (SQE) model are calibrated for silicate materials, based on an extensive training set of representative isolated systems. In total, four calibrations are carried out, two for each model, either using iterative Hirshfeld (HI) charges or ESP grid data computed with Density Functional Theory (DFT) as a reference. Both the static (ground state) reference quantities and their responses to uniform electric fields are included in the fitting procedure. The EEM model fails to describe the response data, while the SQE model quantitatively reproduces all the training data. For the ESP-based parameters, we found that the reference ESP data are only useful at those grid points where the electron density is lower than 10^{-3} a.u. The density value correlates with a distance criterion used for selecting grid points in common ESP fitting schemes. All parameters are validated with DFT computations on an independent set of isolated systems (similar to the training set), and on a set of periodic systems including dense and microporous crystalline silica structures, zirconia, and zirconium silicate. Although the transferability of the parameters to new isolated systems poses no difficulties, the atomic hardness parameters in the HI-based models must be corrected to obtain accurate results for periodic systems. The SQE/ESP model permits the calculation of the ESP with similar accuracy in both isolated and periodic systems.

1 Introduction

The knowledge of the electronic distribution in a molecular system and the variation of the distribution upon a perturbation is a key-point to the understanding of the system behavior at the atomic level. This issue is commonly addressed by quantum-chemical computations, which despite the significant progress in both the computer hardware and computational algorithms, remain an expensive task, especially when studying complex disordered systems. Consequently, significant attention is paid to the development of simplified parametric models that allow the quantities related to the electronic distribution to be computed with a modest cost, while keeping transfer-

ability and/or precision of quantum-chemical calculations. Besides the proper utility, such models can further be used as a part of polarizable force fields, thus permitting realistic studies of large complex systems (yet) inaccessible to quantum-chemical calculations. An additional advantage of parametric models is that they provide much more transparent relation between obtained results and the underlying physics, whereas the outcome of a quantum-mechanical model is harder to interpret due to the complexity of the electronic wavefunction.

The concept of atomic charges provides a tangible means of relating the electronic distribution in the system with its structure and reactivity at the atomic level. Consequently, a wealth of efforts was and is concentrated on the development of models and methods capable of reliable computation of atomic charges and by making use of this information, of getting new insight into the reactivity or structural and dynamical characteristics. Models based on the chemical potential equalization (CPE) or electronegativity equalization (EE) concept appear to be among the most promising candidates to predict the charge distribution in molecules and solids. A number of CPE-based models have been proposed over the past two decades,[1–6] in which atomic charges are variational degrees of freedom and the *electronic* ground state corresponds to an energy minimum under a total charge constraint. The latter is equivalent to the condition that all atomic chemical potentials (or electronegativities) must be equal. This concept was generalized to s- and p-type density basis functions,[7] effectively extending the CPE approach with atomic inducible dipoles. CPE models were applied to different systems such as dense and microporous inorganic solids,[8–10] water,[11–16] ionic liquids,[17] organic liquids,[18–20] other organic molecules,[21–27] biomolecular systems,[28,29] and heterogeneous systems.[30–32] Furthermore, CPE models were also used to study various properties of electronic systems, including exchange-polarization coupling,[13,14] intermolecular charge transfer,[20] charge transfer during bond dissociation,[25,27] non-linear polarizability[15,24] and optical linear response properties such as IR and Rahman intensities.[22]

The earliest CPE model is the Electronegativity Equalization Method (EEM).[1] Although this model was later modified with more realistic electrostatic interactions[2,33] or by adding more interaction sites,[5] EEM and these early extensions predict that the dipole polarizability scales cubicly

with system size, while in the macroscopic limit, one observes a linear relation for dielectric materials.[3,34] The latter weakness is problematic when one tries to apply parameters based on small molecules to larger (dielectric) systems.[35] A few ad-hoc improvements were proposed to overcome this fundamental error.[4,22] We are mainly interested in a recent extension of the EEM, the Split Charge Equilibration (SQE),[4] which fixes this polarizability scaling issue[34] and is capable of properly describing electronic dielectric screening effects in condensed matter.[36] It is especially encouraging that the SQE model exhibits a transition between the EEM-like polarizability scaling for small systems to the linear scaling for extended systems,[34,36] as is often observed in computational studies.[13,37–40]

Since the primary goal of all CPE-based models is the fast and reliable computation of charge distribution in large systems, the parameterization of such a model is done by fitting atomic charges produced by the model to those derived from quantum-chemical calculations. The problem is that the atomic charge is not a quantum-chemical observable[41] and consequently, many electronic density partitioning schemes can be used to produce the reference charge values. The first parametrizations employed atomic charges resulting from the Mulliken population analysis.[1] More recently, atomic charges fitted to the electrostatic potential (ESP) fits,[33,42,43] Natural charges,[26,43] Bader's Atom In Molecule (AIM) charges[29] or charges produced by Stockholder partitioning (Hirshfeld[43] or Hirshfeld-I[26]) were used for this purpose.

Since the beginning of the development of CPE-based models, oxide materials ((alumino)silicates, in particular) were one of important application fields of the corresponding methods.[8–10] A CPE-based model is also an essential component of a polarizable force field for silicates. Nevertheless, to our knowledge, an extensive calibration and validation of the EEM[1] or the SQE[4] for oxide crystals based on a large set of representative cluster models is not yet carried out. (For organic systems, such extensive calibrations can be found in the literature.[19,23,26]) In general, it is not clear yet whether parameters for the EEM or SQE model are simply transferable from small clusters to extended or periodic systems. In this paper, we derive parameters for both models based on DFT computations on a large number of silicate clusters, and use the results to answer the following

research questions:

(i) How well do different CPE-based models reproduce the electronic density distribution, the electrostatic potential, and electronic linear response properties of oxide systems?

(ii) Which types of input data are needed in a reliable parameter calibration protocol for two frequently used CPE models (EEM and SQE)?

(iii) To what extent are the parameters derived from cluster computations transferable to systems of different size and density, in the view of the development of a polarizable force field for condensed matter simulations? Especially the transferability of the parameters to periodic systems is extensively tested.

The paper is organized as follows. Section 2 briefly discusses the EEM and SQE models and presents parameters entering the basic EEM and SQE equations. It goes on to describe the systems used for the calibration and testing of parameters and provides details of quantum-chemical calculations used to obtain reference quantities employed in the parameterization. The final part of the second section deals with the criteria and strategy of the parameter calibration. Results of the calibration of different CPE-based models are reported in the first part of Section 3 which also illustrates the performance of the models in reproducing different characteristics of molecular systems. The second part of the third section deals with the transferability of EEM and SQE parameterizations to periodic systems. The last part of the manuscript provides the conclusions of the work and gives answers to the questions listed in the above paragraph. Additional information that might be of interest for the reader is given in Appendices and as Supplementary Information to the paper.

2 Theoretical models and computational details

2.1 EEM and SQE models

In the framework of CPE models, a general expression for the charge-dependent energy $E(\mathbf{u})$ of a system can be written in a matrix form

$$E(\mathbf{u}) = \mathbf{x}^T\mathbf{u} + \frac{1}{2}\mathbf{u}^T\mathbf{H}\mathbf{u}, \tag{1}$$

where $\mathbf{u}$ is a vector of charge variables, $\mathbf{x}$ a vector with first-order parameters, and $\mathbf{H}$ is the so-called hardness matrix built of second-order parameters and containing information about the electrostatic interactions in the system. Following the variational principle the charge variables u_i can be obtained as a solution of equation

$$\mathbf{u} = -\mathbf{H}^{-1}\mathbf{x}. \tag{2}$$

The dependence of elements of the matrix $\mathbf{H}$ on the interatomic distances makes the charge variables dependent on the geometry and on the environment of atoms in the system.

The first model used in the present study, the electronegativity equalization model by Mortier and co-workers,[1] employs atomic charges, q_i, as the charge variables in Eq. (1). The explicit EEM form for the energy reads

$$E_{\mathrm{EEM}}(\mathbf{q}) = \sum_i \left[\chi_i q_i + \frac{1}{2}\eta_i q_i^2\right] + \frac{1}{2}\sum_i \sum_{j\neq i} q_i q_j J(r_{ji}), \tag{3}$$

where the sums run over atoms in the system, $J(r_{ji})$ is the electrostatic potential between the atoms i and j, and χ_i and η_i are the first- and second-order parameters known as effective atomic electronegativity and hardness, respectively. Minimization of this energy under a total charge constraint yields the equilibrium charge distribution. It is shown that the EEM energy in Eq. (3) can be deduced from the density functional theory[1] and the model is of great help for the prediction of the charge distribution in large systems and for qualitative understanding of system's reactivity

on the basis of reactivity indices. Furthermore, it is often used as an underlying numerical model of the conceptual density functional theory.[44]

In its original formulation the EEM was found to suffer from a number of deficiencies. The most crucial one limiting the use of EEM in the development of polarizable force fields, is the strongly non-linear behavior of the dipole polarizability with the system's size. The SQE model[4] circumvents this problem by employing split-charges as the charge variables in Eq. (1). A split-charge or charge transfer parameter (CTP), p_{ij}, is an amount of charge transfered from atom j to atom i, when the electronegativities of atoms constituting the system get equalized. The CTPs satisfy the following condition

$$p_{ij} = -p_{ji} \tag{4}$$

and the net atomic charge q_i on atom i is then equal

$$q_i = \sum_j^{(i)} p_{ij}, \tag{5}$$

where the superscript (i) is used to denote that the sum runs over atoms j to which the charge transfer from atom i is allowed. It is noteworthy that the condition, Eq. (4), constrains the total charge of system to be equal to zero.

In terms of the split-charges the expression for the energy reads

$$\begin{aligned} E_{\mathrm{SQE}}(\mathbf{p}) = &\sum_i \sum_k^{(i)} \xi_{ik} p_{ik} + \frac{1}{2} \sum_i \sum_k^{(i)} \kappa_{ik} p_{ik}^2 + \frac{1}{2} \sum_i \eta_i \sum_k^{(i)} \sum_l^{(i)} p_{ik} p_{il} \\ &+ \frac{1}{2} \sum_i \sum_{j \neq i} J(r_{ij}) \sum_k^{(i)} \sum_l^{(j)} p_{ik} p_{jl}, \end{aligned} \tag{6}$$

where the sums run over the CTPs and ξ_{ik}, κ_{ik}, and η_i are the first-order and second-order parameters, respectively. The parameters ξ_{ik} and κ_{ik} are the bond electronegativity and bond hardness from the atom-atom charge transfer (AACT) model,[3] respectively. It is worthy of noting that Eq. (6) is given in a redundant set of charge variables p_{ij}. For the practical applications of SQE model one has to write Eq. (6) in the unique set of CTPs using the condition Eq. (4) and the implicit

constraints $\xi_{ji} = -\xi_{ij}$ and $\kappa_{ji} = \kappa_{ij}$.[4]

Although several subtly different forms of the SQE model are possible, it is essential that the bond hardness term (second term in Eq. (6)) is present.[4] In the most general case, when the charge transfers are allowed between all atoms in the system and all the bond hardness parameters are set to zero, the SQE model is isomorphous to the EEM and the parameters in Eq. (6) can then be represented via those of Eq. (3). One can, however, limit the number of CTPs by permitting the charge transfers only between certain atoms. Thus, in the present work only the CTPs for the nearest neighbors (i.e. bonded atoms) are non-zero. Given the parameters in Eq. (6) are known, these CTPs can be found by minimizing the energy $E_{\mathrm{SQE}}(\mathbf{p})$ with respect to the split charges p_{ij}. Eq. (3) and Eq. (6) give the charge-dependent energy in the EEM and SQE representations, respectively, for isolated systems. The modification of the equations as well as of the hardness matrix elements for periodic systems are given in Appendix A. It is worthy of note that the condition of chemical potential (electronegativity) equalization never appears in the SQE equations because the total charge is implicitly constrained to be zero. It can, however, be shown that the minimization of energy, Eq. (6), with respect to CTPs is equivalent to the use of electronegativity equalization principle in the EEM.[4]

In Eq. (3) and Eq. (6) $J(r_{ij})$ stands for an electrostatic interaction potential that is taken as Coulombic potential $1/r_{ij}$ between the two point charges in the simplest case. However, a realistic electrostatic interaction between a pair of atoms significantly deviates from the $1/r$ dependence at short distances. Hence, to mimic the interactions between bonded atoms, it is more realistic to use a potential $J(r_{ij})$ that describes the interaction energy between distributed charges.[2,33] In the present work the $J(r_{ij})$ potential function was taken in the form

$$J(r_{ij}) = \frac{\mathrm{erf}(\alpha_{ij} r_{ij})}{r_{ij}}, \tag{7}$$

which corresponds to the interaction potential between two Gaussian charge distributions $g(\mathbf{r})$

given by

$$g_i(\mathbf{r}) = \left(\frac{1}{2\pi R_i^2}\right)^{3/2} \exp(-(\mathbf{r}-\mathbf{r}_i)^2/2R_i^2), \tag{8}$$

where $\mathbf{r}_i$ is the position of atom i and the parameter R_i (the standard deviation) can be viewed as an effective radius of atom i. With this expression the parameter α_{ij} in Eq. (7) is obtained as

$$\alpha_{ij} = \left(\frac{1}{2R_i^2 + 2R_j^2}\right)^{1/2}. \tag{9}$$

The quantities χ_i, η_i, ξ_{ik}, κ_{ik}, and R_i in Eq. (3), Eq. (6), and Eq. (8) are the model parameters whose values were obtained in a fitting procedure as described below.

2.2 Computational schemes

The EEM and SQE models are calibrated on two different reference quantities derived from quantum-chemical calculations. The first calibration scheme follows a common way of parametrizing CPE models and uses atomic charges as reference quantities to obtain the model parameters. In the present work the reference charges are computed with the iterative Hirshfeld (HI) method.[45] Since the atomic charge is not an observable, the choice of population analysis scheme is not unique and other charge schemes could have been used.[42,43] The choice of the HI partitioning scheme is based (i) on the fact that the computation of HI charges relies on the partitioning of the electronic density, i.e. a quantum-chemical observable, (ii) on a weak dependence of the HI charges on the basis set used in the quantum-chemical calculations,[46] (iii) on the quality of the ESP predicted by HI charges,[47] and (iv) on the robustness of HI charges with respect to conformational changes.[35,48] The second parameterization scheme is based on the reference values of the electrostatic potential (ESP) in the view of the application of the EEM and SQE models to the development of polarizable force fields. The combination of model/quantity finally resulted in four computational schemes employed in the parameterization EEM/HI, EEM/ESP, SQE/HI, SQE/ESP. The reference data used for the parameter calibration are derived from quantum-chemical calcula-

tions on isolated molecules. Following procedures outlined below, the isolated systems are divided into training and validation sets that allowed assessment of the transferability of parameters: the training set is used for the actual calibration, while the validation set is used to test how well the parameters would work for molecules not present in the training set. In addition to the validation set containing isolated molecules, quantum-mechanical computations on periodic systems are used to test the transferability to crystalline oxides.

2.3 Training and validation sets

The database of isolated molecule structures consists of 207 oxide clusters containing three-coordinated aluminum cations, and four-coordinated silicon and zirconium cations. The finite clusters are mostly hydrogen-terminated. Some hydroxyl terminations are present, but none of them form internal hydrogen bonds that distort the cluster geometry. The chemical formula of these structures can be written as $Si_xAl_yZr_zO_{\frac{1}{2}(4x+3y+4z-n-m)}H_n(OH)_m$. The molecular structures are distributed over a training set (103 molecules) and validation set (104 molecules), such that they are statistically representative for each other. Table 1 gathers the relevant information on each set. The list and the geometry of clusters in the training and validation sets are available as Supporting Information.

The database of periodic systems contains crystalline modifications of silicon oxide, tetragonal modification of zirconia (t-ZrO_2), and zirconium silicate ($ZrSiO_4$, zircon polymorph). Silica structures chosen for these calculations were α-quartz, α-cristobalite, and four all-silica zeolite structures of the JBW, DFT, SOD, and NPO topologies.[49] This selection of periodic systems, which are not used for the calibration of the parameters, permits to test the transferability of parameters on both dense and microporous materials. Table 2 provides some pertinent characteristics of the structures; CIF files and images of the optimized structures can be found in Supporting Information.

2.4 Computational details

Isolated Molecule calculations The calculations on the isolated systems were performed at the DFT level with the B3LYP exchange-correlation functional.[50] The 6-311+(d,p)[51,52] all-electron basis set was used for the H, O, Al, and Si atoms, while the Zr atoms were described with the LANL2DZ effective core potentials.[53] The geometry of molecules was first optimized without any symmetry constraints and was followed by the calculation of the electronic density and electrostatic potential values on a three-dimensional grid of points (cubes files). The calculations were done with Gaussian03 code.[54]

Periodic calculations The calculations of periodic systems were carried at the same level as the isolated molecule calculations, i.e. within DFT using the B3LYP[50] functional, and they were done with the CRYSTAL06 code.[55] Atoms were described by all-electron basis sets: 976-31d621G for Zr atoms, 86-311G* for Si atoms, 85-11G* for Al atoms, 6-31G* for O atoms and 8-21G* for H atoms.[56–59] The Brillouin zone integration was carried out over a Monkhorst-Pack grid of $6 \times 6 \times 6$ k-points.[60] Model structures were optimized under the constraints imposed by the symmetry of crystalline lattices. These structural optimizations were followed by the computation of cube files with the electronic density and electrostatic potential.

Computation of Hirshfeld-I charges The computation of the Hirshfeld-I charges is carried out with HiPart.[61] For the isolated molecules, the Hirshfeld-I charges are derived using the procedure outlined in the work of Bultinck *et al.*[45] For the periodic systems, a slightly modified procedure is used for technical reasons. Due to the presence of the cusps in the electron density, one should use spherical atom-centered grids to carry out the numerical integrations required for the Hirshfeld-I partitioning.[62] For the isolated systems, this approach is feasible because the formatted checkpoint files from Gaussian03 provide sufficient information to evaluate the electron density on arbitrary grids. However, it is technically infeasible to extract similar grid data from CRYSTAL06 computations in a post-processing analysis. Therefore we used the cubic grids from the density cube

files to approximate the Hirshfeld-I charges as follows: (i) first the contribution to the density from the core orbitals is subtracted, using spherically averaged core-densities from atomic computations, and (ii) the remainder (which is a smooth function that can be integrated on cubic grids) is partitioned using Hirshfeld-I algorithm.

2.5 Parametrization strategy

Cost functions In line with the calibration of EEM and SQE parameters for organic systems in earlier work,[26] two types of cost functions (X) are used in the calibration procedure: static (X_{S}) and response (X_{R}) ones. The former is based on the values of the reference quantity itself (HI charge or ESP), whereas the latter corresponds to the derivatives of the values with respect to an external perturbation taken in the form of an uniform electric field. For each of the four computational schemes (EEM/HI, etc.) a series of parameters is calibrated with the cost function

$$X_\lambda = (X_{\mathrm{S}} + \lambda X_{\mathrm{R}})/(1+\lambda) \tag{10}$$

where the weight λ is scanned over several orders of magnitude. When λ approaches zero, the total cost approaches the static cost (X_{S}), while the response cost has a minimal effect. The lower bound of the λ-scan was chosen such that the static cost function converges to a constant value. Similar considerations were used to determine the upper bound of the λ-scan. Some testing revealed that the transition from a purely static cost function to a pure response cost function is found for λ going from 10^{-6} to 10^{0}. The final choice of λ (after performing the scans) is based on the following criteria:

- Small values of both the X_{S} and X_{R} cost functions,
- Minimum number of constraints (*vide infra*),
- Small value of the condition number of the Hessian matrix of the cost function in Eq. (10). This guarantees that the parameters have a minimal sensitivity to "noise" in the reference characteristics.[63]

The static X_S and response X_R cost functions for each CPE model (EEM or SQE) and for each reference quantity (HI charges or ESP) are discussed in the Appendix B in detail.

Given that the response cost function (X_R) only depends on the second-order parameters, one could try to reduce the number of independent parameters by splitting the calibration into two steps: (i) first determine the second-order parameters that minimize X_R, (ii) consequently keep these second-order parameters fixed and find the remaimder of the parameters by minimizing the static cost function (X_S). However, this was not helpful because the second-order parameters found in the first step were systematically poorly determined and reached unreasonable values, making it impossible to find useful first-order parameters to complete the model. Only when a linear combination of both cost functions was used, the calibrations became well-conditioned and reasonable parameters could be obtained.

Constraints There are a series of inequality constraints for the second-order parameters to keep the matrix of second order coefficient in both the EEM and SQE positive definite:[26,34]

- The radii of the charge distributions have to be larger than 0.1 Bohr; we do not take zero as the lower bound due to the next inequality constraint.
- The atomic hardness has to be larger than the self-interaction potential of atomic Gaussian charge distribution: $\eta_i \geq (\sqrt{\pi}R_i)^{-1}$
- The bond hardness κ_{ik} has to be positive.

Two additional constraints on the parameters must be introduced because the cost function, Eq. (10), is not sensitive to deviations from these two constraints.[63] Thus, in case of EEM the effective electronegativity of hydrogen atom is fixed at $\chi^*_H = 0.0$ eV. In case of the SQE model the effective atomic hardness of hydrogen is kept at the lower bound corresponding the radius R_i of the charge distribution in Eq. (8), i. e. $\eta^*_H = (\sqrt{\pi}R_H)^{-1}$.

Minimizer algorithm Due to the non-linear dependence of the EEM or SQE charges on the parameters, the calibration is a non-linear least-squares (NLLSQ) problem. A conjugate gradi-

ent optimizer with a diagonal preconditioner is used to optimize the parameters.[64] An active set algorithm is used to implement the inequality constraints. A inequality constraint only becomes active when the optimizer tries to push the parameters into the infeasible region. As soon as the dot product of the gradient of the cost function and the normal of a constraint becomes negative, the corresponding constraint is deactivated. The initial values for the parameters χ_i^*, η_i^*, ξ_{ik}^*, κ_{ik}^*, and R_i, are 0 eV, 20 eV, 0 eV, 5 eV and 1 Å, respectively. The optimal parameters are not sensitive to the initial values. For parameters that have a lower bound, the initial values are set such that all inequality constraints are satisfied with a considerable margin.

3 Results and discussion

3.1 Choice of the best set of parameters

In order to find an optimal set of parameters for each computational scheme (model/quantity combination) the coefficient λ in Eq. (10) was varied over several orders of magnitude in the $[0, 1]$ interval and a set of parameters minimizing X_λ was found in a NLLSQ fitting procedure for each value of λ. The best set of parameters was then chosen using the above criteria, i.e. the values of the static and response cost functions, the condition number, and the number of constraints.

Figure 1 presents the dependence of the total (X_λ), static (X_S), and response (X_R) cost functions (left panel), the number of constraints, and the condition number (right panel) on the weight λ for the EEM/HI scheme. The blue and red horizontal lines indicate "worst-case" values of the static and response cost functions, respectively. This "worst-case" value corresponds to the value of the cost function when the model would predict zero values for the target quantities, i.e. when there are no electrostatic interactions or the charges are not sensitive to an external electric field. In principle, the cost function may become higher when the predicted values have the wrong sign. However, in the latter case it is more "accurate" not to model electrostatic interactions at all, i.e. use the "worst-case" limit. Therefore the "worst-case" values of the cost function indicate a threshold below which the use of the corresponding quantity in the calibration procedure starts to make

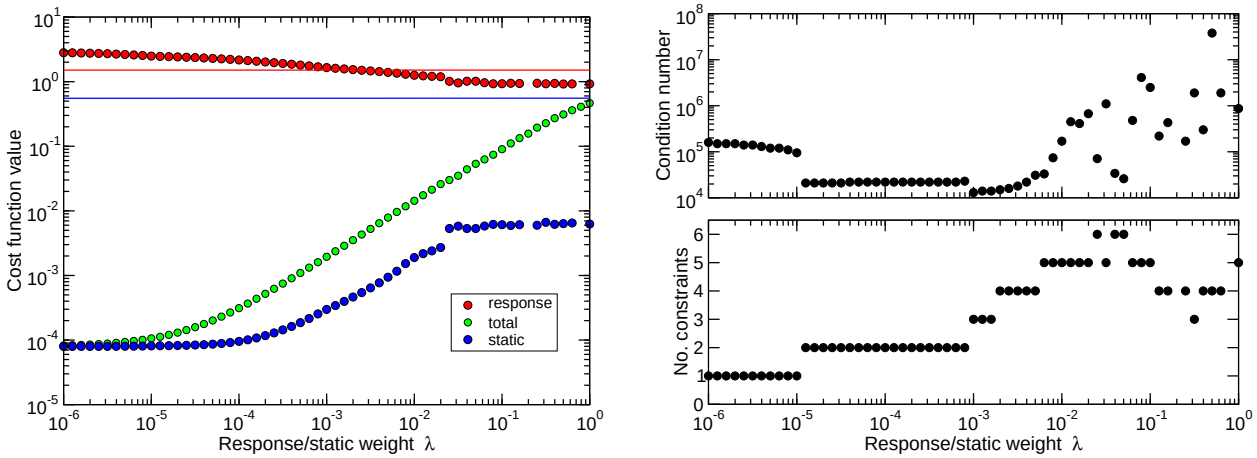

Figure 1: Dependence of EEM/HI total X_λ, static X_S, and response X_R cost functions on the weight λ, Eq. (10). Horizontal blue and red lines indicate "worst-case" values of static and response cost functions, respectively (see text for discussion).

sense.

Figure 1 shows that at low values of λ (pure static case) the response cost function is above the threshold level. An increase of λ does not lead to an improvement of the situation: when the response cost function drops below the "worst-case" value, the static cost function has already increased by more than an order of magnitude. At even larger values of λ, the total cost function becomes ill-defined and a large number of constraints needs to be used to get the minimization of Eq. (10) converged (see Figure 1, right panel). It is also noteworthy that regardless the λ value the response cost function remains close to the "worst-case" value. Such a behavior points to the fact that the EEM with HI charges only makes sense when the static cost function is used. No improvement is possible by including response information. The EEM/ESP scheme reveals similar behavior. (Figure of λ-scan is included in the Supporting Information.) The main difference between the EEM/HI and the EEM/ESP calibrations, is that in the latter case the optimal value of the static cost function is merely 0.18 times the "worst case" value, while for EEM/HI this ratio is as low as 1.4×10^{-4}. Because of this poor result, we do not consider the EEM/ESP model for the validation in the remainder of the paper.

The above results allow to conclude that the electronegativity equalization model is only use-

ful for computing static characteristics, such as the charge distribution and (to lesser extent) the electrostatic potential, whereas it fails to mimic the response of these quantities to an applied external perturbation. Consequently, sets of parameters corresponding to pure static cost functions in Eq. (10) were chosen for the EEM-based schemes.

The SQE-based schemes behave differently. The best parameter set for the SQE/HI scheme can be found at $\lambda = 10^{-3.4}$, where both the static and response cost functions increase by 60 % above their minimum values, whereas the total cost function is still close to its optimum value. (Figure of the λ-scan is included in the Supporting Information.) It should also be noted that, in contrast to the EEM/HI scheme, the parameters obtained in the limit of high values of λ can be useful because both the partial cost functions remain significantly below their respective "worst-case" values, while a small value of the condition number points to the stability of the NLLSQ solution with respect to "noise" in the training set.

Figure 2 displays the values of the total, static, and response cost functions (left panel), the number of constraints and value of the condition number (right panel) as a function of the weight λ in Eq. (10) for the SQE/ESP scheme. Again, in contrast to the EEM/HI model (Figure 1), both the static and response SQE/ESP cost functions lie always below their respective "worst-case" values and a compromise between the static and response characteristics can be found for the value of $\lambda = 10^{-5}$ with only one half-open constraint active ($\eta_{\mathrm{O}} = (\sqrt{\pi}R_{\mathrm{O}})^{-1}$) and a small value of the condition number. Note that an increase of λ value leads to errors in the static characteristics that cannot be compensated by the decrease of the error in the response data. Furthermore a larger number of constraints needs to be used to converge the function, Eq. (10), to the minimum.

We can conclude that the SQE parameters can be calibrated to reproduce both the static and response data, either using atomic charges or the ESP as target data. This is a distinct improvement compared to the EEM where only the EEM/HI calibration in the static limit behaves satisfactory.

Table 3 gathers the best sets of EEM and SQE model parameters that were chosen according to the criteria mentioned above after the analysis of each of four model/quantity combinations.

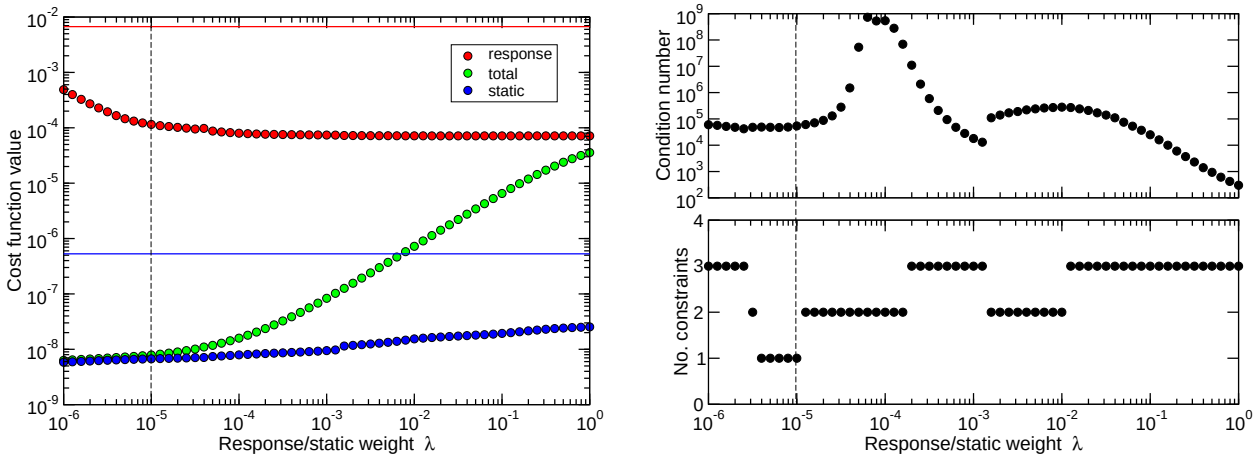

Figure 2: Dependence of SQE/ESP total X_λ, static X_S, and response X_R cost functions on weight λ, Eq. (10). Horizontal blue and red lines indicate "worst-case" values of static and response cost functions, respectively (see text for discussion; vertical dashed lines indicate the selected point.

3.2 Charge calculation with EEM and SQE models

Isolated systems. Despite the ambiguity of the atomic charge and the existence of many charge partitioning schemes, atomic charges provide an extremely useful guidance for the understanding of many properties of system at the atomic level. Consequently, models permitting a fast and reliable computation of charge distribution are of significant interest. The EEM was such a scheme from the very beginning of its development.

Figure 3 shows the correlation of charges computed by the EEM or SQE models with the reference iterative Hirshfeld charges derived from the results of DFT calculations on molecules of the validation set. Both models perform remarkably well. The root mean square deviation (RMSD) between the SQE/HI and DFT/HI charges, 0.0213 e, is slightly lower than the RMSD between the EEM/HI and DFT/HI charges, 0.0284 e. These relatively small RMSD values testify the transferability of both the CPE models for the computation of charge distribution in isolated systems of modest (up to 20-30 atoms) size.

Periodic systems. In what follows, we will validate to what extent the EEM/HI and SQE/HI calibrations can reliably predict the atomic charges in periodic systems. After all, the parameters

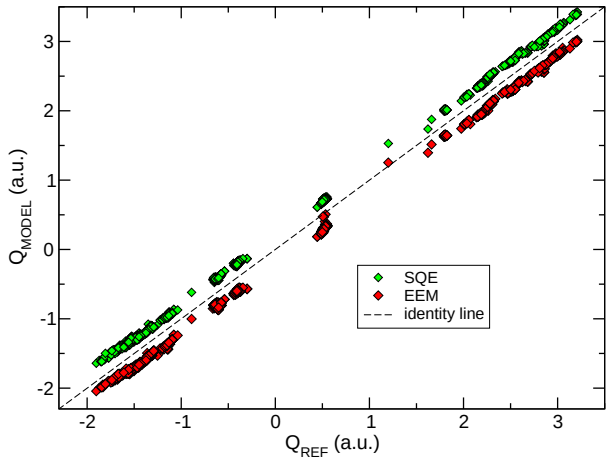

Figure 3: Correlation between the reference DFT HI charges and those predicted by the EEM and SQE models for molecules in the validation set. For the sake of clarity the EEM and SQE data were equally shifted down and up, respectively, along the *y*-axis.

are based on a training set containing only isolated systems. The left-hand part of Table 4 reports the atomic charges computed with EEM/HI and SQE/HI schemes for the periodic structures of the validation set and compares the charges with those obtained by applying the HI procedure to the electronic density from periodic DFT computations (DFT/HI column). The performance of the schemes in the prediction of the charge distribution in periodic systems is not that good as one might expect. One sees that the semi-empirical models largely overestimate charge transfer from metallic cations to the oxygen atoms making the systems more ionic than they are according to the results of iterative Hirshfeld analysis of quantum-chemical data.

Results of the SQE/HI model for the zirconium silicate and zirconia are worth a special remark. The model yields a charge for the Zr cation in the zirconium silicate that is larger than the formal ionic charge. The reason for such an artifact is the following. The Zr cations in the structure are eight-fold coordinated with four O atoms at a distance of 2.163 Å and four O atoms at a distance of 2.287 Å, whereas the SQE/HI parameters were calibrated on isolated systems having four-fold coordinated Zr cations. Consequently, the use of the bonding-specific parameters together with the explicit charge-transfer channels in the SQE scheme leads to an overestimation of the charge flow from the zirconium to oxygen atoms. In the case of zirconia the cations have four oxygens

at a distance of 2.096 Å and four at 2.371 Å. The coordination number of the cations used in the SQE calculation was then taken to be four and the computed value of Zr atomic charge is below the formal ionic charge. Allowing the charge transfer also to the four next-nearest oxygens at the 2.371 Å distance leads to the zirconium atoms charge $q_{Zr} = 4.3212\ |e|$, that is, like in the case of $ZrSiO_4$, larger than the formal ionic charge of the cation. Note that the problem does not occur with the EEM scheme because it has only atom-based first-order terms in Eq. (3). A way of improving the SQE model is the use of distance-dependent bond electronegativities $\xi_{ji}(r_{ji})$, as it was suggested by Chen and Martínez.[6] Indeed, an attempt of introducing such a parameter for the Zr-O bond in the form

$$\xi_{ij}(r_{ij}) = \frac{\xi_{ij}}{1+\exp(a(r_{ij}-b))} \tag{11}$$

with *ad hoc* parameters $a = 18.0$ Å^{-1} and $b = 2.38$ Å resulted in a decrease of charge of eight-fold coordinated Zr atoms in $ZrSiO_4$ to $q_{Zr} = 3.9821\ |e|$ without a significant change of charge of the four-fold coordinated zirconium in the tetragonal zirconia.

For the crystalline silica polymorphs both semi-empirical schemes give similar precision in the estimation of Hirshfeld-I charges in the periodic systems with the mean relative errors of 27 % and 22 % for the EEM/HI and SQE/HI models, respectively. It is not immediately clear why the transferability of the EEM/HI and SQE/HI to the periodic systems is only qualitative and not quantitative. There are several plausible explanations, some of which listed below:

1. The B3LYP XC functional in Gaussian03 package is slightly different from the B3LYP used in CRYSTAL06 code: the programs employ VWN3 and VWN5 LDA correlation functionals, respectively. However, we think that the difference can hardly account for the discrepancy between the semi-empirical and DFT HI charges.

2. Different basis sets are used for the periodic and cluster computations. Given a weak dependence of Hirshfeld-I charges on the basis set, this effect is expected to be small, although it may still contribute to the discrepancy.

3. The molecules in the training set are yet too small to capture the characteristics of atoms in

the solid state. This explanation is the most plausible, as we show below, despite the fact that no notable sensitivity of the iterative Hirshfeld charges to the size of isolated molecules can be inferred from Figure 3. The difference between the reference and model charges are so small that it is impossible to deduce some dependence of the errors on molecule size.

Corrections for periodic systems. Below we show how the parameters derived from the isolated systems in the training set can be corrected to accurately reproduce the DFT/HI charges in condensed phase, and how this correction can be rationalized. Let us consider the EEM model and write two equations (cf. Eq. (2))

$$\mathbf{q}_{\mathrm{EEM}} = -\mathbf{H}^{-1}\mathbf{x} \tag{12}$$

$$\mathbf{q}_{\mathrm{DFT/HI}} = -\mathbf{H}^{*-1}\mathbf{x}^{*}, \tag{13}$$

where we introduced $\mathbf{H}^{*}$ and $\mathbf{x}^{*}$ as (hypothetical) more accurate descriptions of the hardness matrix and the electronegativity vector, respectively, in periodic systems, which result exactly in the DFT Hirshfeld-I charges. The last equation can be rewritten so that the $\mathbf{H}^{*}$ matrix also includes the difference between the $\mathbf{x}$ and $\mathbf{x}^{*}$ vectors,[63] i.e.

$$\mathbf{q}_{\mathrm{DFT/HI}} = -\mathbf{H}^{*-1}\mathbf{x}. \tag{14}$$

It needs to be validated how the hardness matrix needs to be adapted to get more accurate predictions for the charges in periodic systems. Let us first assume that the main difference between the $\mathbf{H}$ and $\mathbf{H}^{*}$ matrices is due to the diagonal elements. We will further test if this assumption is consistent with the observed errors. Eq. (12) and Eq. (13) then become

$$\mathbf{q}_{\mathrm{EEM}} = -\mathbf{H}^{-1}\mathbf{x} \tag{15}$$

$$\mathbf{q}_{\mathrm{DFT/HI}} = -(\mathbf{H}+\mathbf{G})^{-1}\mathbf{x}, \tag{16}$$

where **G** is a diagonal matrix of corrections to the hardness matrix that should reduce the observed errors. Straightforward elimination of **x** from both equations can be used to estimate the corrections to the diagonal elements of the EEM hardness matrix **H** :

$$\mathbf{H}(\mathbf{q}_{\mathrm{EEM}} - \mathbf{q}_{\mathrm{DFT/HI}}) = \mathbf{G}\mathbf{q}_{\mathrm{DFT/HI}}. \tag{17}$$

Table 5 reports elements of the **G** matrix for the Si and O atoms in some of silica structures of Table 2. Clearly, the values are structure dependent, but the difference is not very large and we can compute mean values which are $G_O = 0.7668$ (0.0958) eV and $G_{Si} = 0.4216$ (0.0354) eV. Making use of these values, we now calculate the HI charges of atoms in all our silica polymorphs by modifying the diagonal elements of the hardness matrix. The results are gathered in Table 6 that shows a nice performance of EEM/HI model modified for the solid state computations ($\mathrm{EEM_S}$/HI column): the mean relative error for the charges on oxygens drops from 27 % to 1 %. Interestingly, the use of the same G_{ii} parameters in the SQE/HI scheme leads to similar spectacular improvement of HI charges whose mean relative deviation from the reference DFT values decreases to 1 % (Table 6, $\mathrm{SQE_S}$/HI column).

The above results can be rationalized in the following way. Table 5 shows that G_{ii} parameters are *positive*, which represents an increase of the effective hardness of atoms in solids as compared to atoms in molecules. The inverse of the hardness is the softness, which is related to the polarizability.[44,65] Therefore the increase of atomic hardness needed to mimic HI charges in solids, can be interpreted as an evidence of *decrease of polarizability* of atoms, when going from molecules to solids. Indeed, the crystalline field in solids confines the electrons, which then have a reduced ability to respond to an external perturbation in comparison to a molecule. A generalization of the above correction scheme can provide a way of obtaining EEM and SQE charge-based models transferable among a large palette of systems of different size and densities.

Although this correction scheme is very effective, it is clear that this problem must be analyzed in more detail in future work. For example, it is not yet clear how this correction scales from

zero to the values given above when going from the isolated molecules in the training set to infinite periodic systems, i.e. starting from which size the system cannot be considered anymore as isolated.

3.3 Electrostatic potential

Isolated systems As it was mentioned above, the EEM/ESP scheme is useful only for computing the static characteristics and therefore, the following presentation and discussion are focused on the ESP calibrated SQE model that is capable of modeling both static and response ESP values.

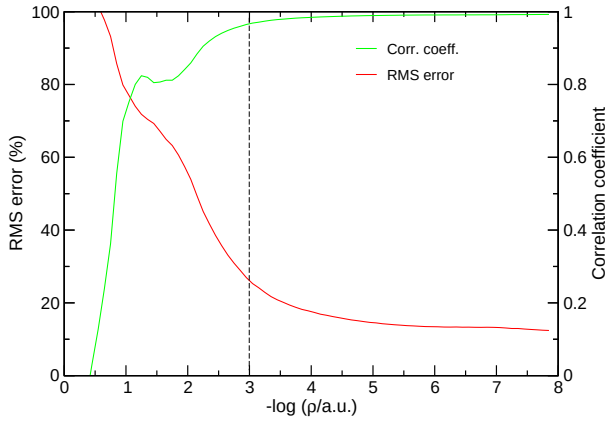

Figure 4: Relative error in ESP values and correlation coefficient between reference and model ESP values *vs* electronic density for molecules in validation set; results obtained with SQE/ESP parameters of Table 3. Vertical dashed line corresponds to the threshold density value.

Initially we investigate how well the SQE/ESP model reproduces the reference ESP values from the DFT computations on the isolated systems. Figure 4 shows the relative RMS error and the correlation coefficient between the SQE/ESP predictions and DFT/ESP reference data as function of the electronic density computed in the same point. One sees that the model reproduces the ESP around the molecules with the minimum mean error of ca 13 % and with the value of 0.99 for correlation coefficient. One observes a significant increase of the error accompanied by a decrease of the correlation coefficient for the values of the density above 10^{-3} a.u. This value can therefore be taken as a threshold beyond which the results obtained with the SQE/ESP model become unreliable. These results can be understood by analyzing the Poisson equation (in atomic

units),

$$\nabla^2 V(\mathbf{r}) = -4\pi\rho_{\text{total}}(\mathbf{r}), \tag{18}$$

where V is the electrostatic potential and ρ_{total} represents the *complete* charge density. In the case of a DFT computation, ρ_{total} consists of the nuclear plus electronic charge density. In the charge model used for the ESP fit, ρ_{total} just contains the Gaussian charge densities from Eq. (8). Only at the points $\mathbf{r}$ where the right-hand side of the Poisson equation is almost the same in both the DFT description and the charge model, one can expect the charge model to work. This condition is fulfilled at those distances from the nuclei, where the nuclear charge screened by the electronic density is well approximated by the spherical Gaussian charge distribution, Eq. (8), i.e. sufficiently far from the nuclei. For that reason, one can only use grid points "outside" the molecule, when computing ESP-fitted charges.

Singh and Kollman[66] found that grid points for fitting ESP charges need to be chosen at shells of at least 1.2 times of atomic Van der Waals radii, which is a quick method for selecting points where the electron density is low. The correlation between the distance from the oxygen atoms and the electron density in a given point is shown in Figure 5. This plot is based on data from all molecules in the training and validation set. Interestingly, the threshold density value of 10^{-3} roughly corresponds to the distance of 1.8 Å that is ca. 1.2 times the Van der Waals radius of an O atom,[67] in agreement with the finding by Kollman *et al.*[66] The spread of the electron density for a given distance, however, shows that it is safer to rely on the actual electron density (instead of fixed radii) to determine the grid points for the ESP fitting procedure. It is also questionable to what extent such radii are transferable between different oxidation states of a given element. The volume of an atom, and hence also its radius, depends on the population, which can only be deduced once the charges are fitted.

It is also of interest to test the ability of the SQE/ESP model to mimic other properties than the ones used in the calibration procedure, i.e charges and the electrostatic potentials. Two such properties were chosen: the dipole moment and the dipole polarizability.[7] Figure 6 and Figure 7 present the comparison of the dipole moment components and of the principal components of the dipole

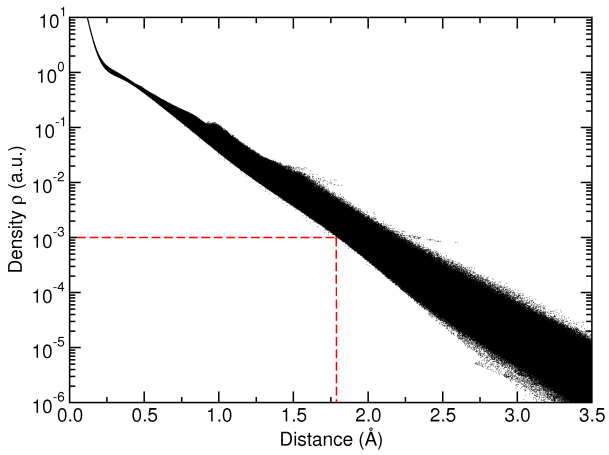

Figure 5: Plot of electronic density value in a given point as a function of distance from closest oxygen atom. Dashed lines indicate coordinates for $\rho = 10^{-3}$ a.u.

polarizability tensor computed with SQE/ESP model with those obtained in the quantum-chemical calculations for molecules of the validation set. The agreement between the model predictions and the reference data is very satisfactory. Again, the current SQE/ESP model uses spherically symmetric atomic density basis and consequently, the model is not capable of predicting the out-of-plane polarizability component of plane molecules (points on the x-axis in Figure 7).

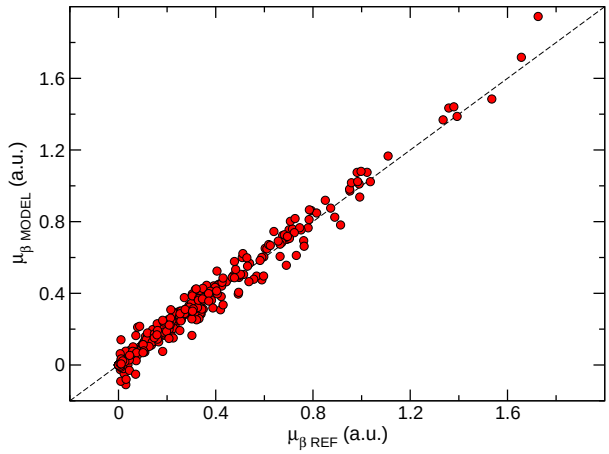

Figure 6: Comparison of the dipole moment components for the molecules in the validation set, obtained with the reference DFT method and with the SQE/ESP model.

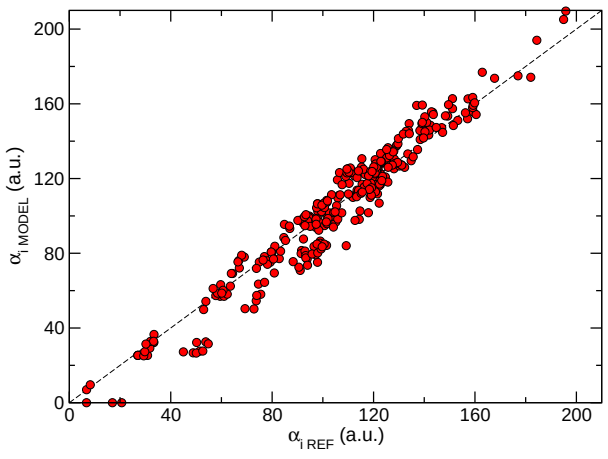

Figure 7: Comparison of all principal components α_i of the polarizability tensors for the molecules in the validation set, obtained with the reference DFT method and with the SQE/ESP model.

Periodic models. The transferability of the SQE/ESP parameters to periodic systems was tested by comparing the electrostatic potential on a cubic grid computed with the SQE/ESP model and the DFT computations for the structures listed in Table 2. According to the results reported above, the ESP values should be comparable in those points of the 3D grid where the value of the electronic density does not exceed the threshold 10^{-3} a.u. This issue, however, complicates the comparison in dense structures because only a relatively small number of points fulfill the criterion. Figure 8 (left pannel) shows those volumes in the crystallographic unit cell of α-quartz structure, where the value of the electronic density is below the threshold value $\rho = 10^{-3}$ a.u. One sees that only a small part of the volume (ca. 5 %) can be used to compute the values of the ESP for the structure. Consequently, all dense structures (t-ZrO_2, $ZrSiO_4$, α-cristiobalite) provide a hard test for the predictive power of the model. The situation is markedly different for zeolitic structures. Thus, the electronic density is below the threshold value in ca. 36 % of grid points in the sodalite structure (Figure 8, right pannel). Therefore, to asses the transferability of the SQE/ESP model we have chosen to present two extreme cases of the most and least dense structures: α-quartz and sodalite (SOD), respectively (cf. Table Table 2).

Figure 9 and Figure 10 present the relative error and the correlation coefficient for the ESP values computed as a function of the electronic density in the α-quartz and sodalite structures.

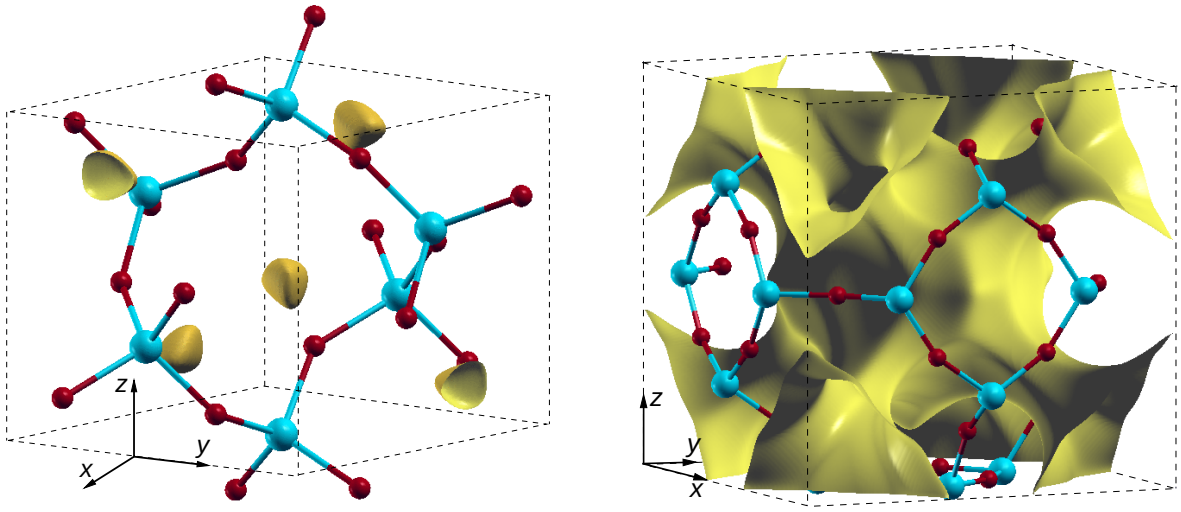

Figure 8: Left: image of α-quartz structure showing the region, where electronic density value is below 10^{-3} a.u. (in yellow color); silicon and oxygen atoms are shown as cyan and red balls, respectively. Right: image of sodalite unit cell with the isodensity surface corresponding to $\rho = 10^{-3}$ a.u. The images were generated with XCrySDen program.[68]

Despite the difference in the number of points employed in the electrostatic potential calculations, the ESP is reproduced in both structures equally well. The relative error does not exceed 20 % with the correlation coefficient above 0.98, at the grid points, where the electronic density value remains below the threshold 10^{-3}. These confidence values are close to those obtained for isolated molecules and indicate a good transferability of the SQE/ESP parameters from isolated to periodic systems.

Figure 11 presents the ESP and the ESP gradient obtained for the sodalite structure along the <100> direction (lines connecting the centers of opposite 4R rings) in the DFT computation and with the SQE/ESP model. The agreement between the semi-empirical and *ab initio* data is remarkable. It is noteworthy that the electronic density in the regions with the coordinate less than 1 Å and more than 8 Å is above the threshold value 10^{-3} and thus, the SQE results in this regions can not be considered as reliable.

Because the SQE/ESP model successfully reproduces the ESP derived from the DFT computations, one would expect that the SQE/ESP charges correlate with ESP fitted charges from the

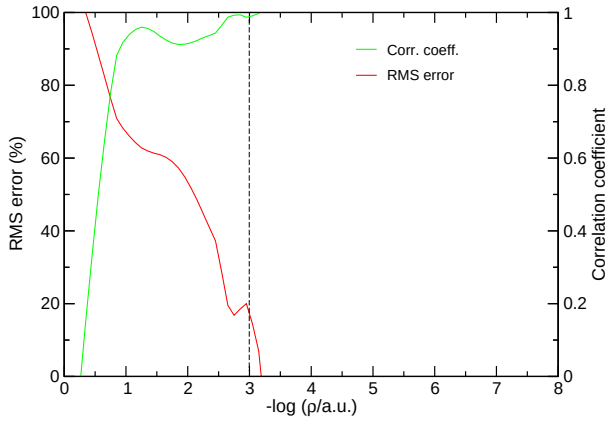

Figure 9: Relative error in ESP values and correlation coefficient between reference and model ESP values *vs* electronic density in α-quartz structure; results obtained with SQE/ESP parameters of Table 3. Vertical dashed line corresponds to the threshold density value $\rho = 10^{-3}$ a.u.

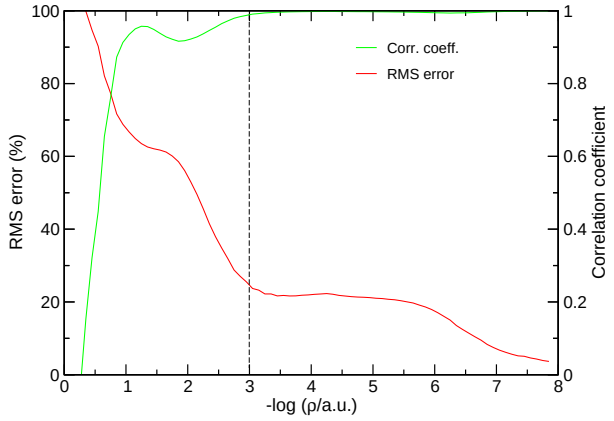

Figure 10: Relative error in ESP values and correlation coefficient between reference and model ESP values *vs* electronic density in sodalite structure; results obtained with SQE/ESP parameters of Table 3. Vertical dashed line corresponds to the threshold density value $\rho = 10^{-3}$ a.u.

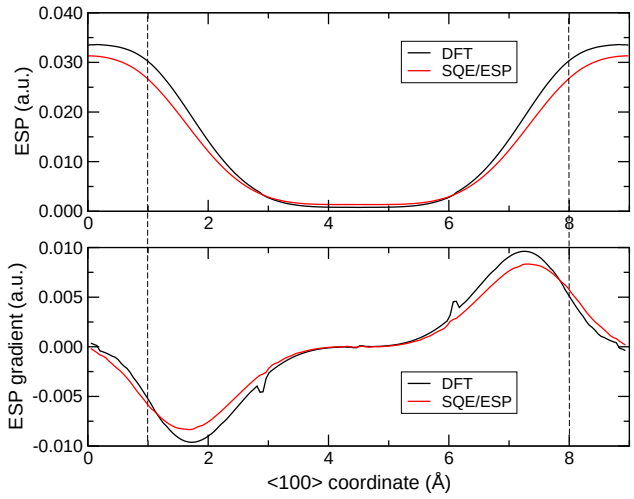

Figure 11: ESP (upper panel) and ESP gradient (lower panel) obtained in periodic DFT and SQE/ESP computations along the <100> direction in sodalite structure (lines connecting the centers of opposite 4R rings). The electron density in the region between the dashed lines is below the threshold value $\rho = 10^{-3}$ a.u.

DFT computations. The right-hand part of Table 4 reports ESP charges for the periodic structures predicted with the EEM/ESP and SQE/ESP schemes and compares the values with ESP charges derived from the DFT calculations. The EEM model generally overestimates, while the SQE one underestimates the ESP charges. It is notable that both models do reproduce the trends in the magnitude of charges from one structure to another. There are, however, a few difficult cases. Thus, the variation of ESP charges predicted by SQE model between the atoms in the α-quartz and α-cristobalite structures is significantly smaller than that obtained from the DFT results. Furthermore, one can note that the EEM scheme fails to reproduce the ESP charge distribution in the zircon polymorph of $ZrSiO_4$: the EEM/ESP charge on the Si atom is larger than that on Zr one, whereas the DFT/ESP charges show the inverse trend. Morover, the decrease of the oxygen charge as compared to the pure silica is absent in the EEM/ESP model, while the SQE/ESP predicts this trend qualitatively.

There are large discrepancies in the comparison of ESP fitted charges and the corresponding values computed with the SQE/ESP and EEM/ESP models. However, such differences do not mean that the results of ESP-based models are unreliable. It is well known that large changes in

atomic charges can lead to only small changes in the ESP surrounding the atoms. This also means that two sets of charges that give a fair reproduction of the ESP may be manifestly different.[69] This lack of sensitivity of ESP fitted charges results in unpredictable contributions, i.e. apparent as noise on the charges, which makes it hard to use them for a direct comparison.[35] The right-hand part of Table 4 contains three sets of charges that attempt to give a good description of the electrostatic potential in the interstitial regions of crystalline structures, but that does not imply these charges should be equal or show the same subtle trends.

Recently, Campanã and co-workers proposed a method for generating ESP charges for periodic systems from results of periodic-quantum-chemical calculations.[70] The authors noted that in the calculations employing the plane-wave basis set and pseudo-potentials, the ESP values are defined up to a constant offset and thus the cost function should be based on the difference of potential values in the grid points rather than on the values themselves. Making use of their approach and of different plane-wave codes (VASP, CPMD, and SIESTA), Campanã *et al.* reported ESP charges for the Si atoms of sodalite structure in the range 1.151 to 1.389 $|e|$. These values have the same order of magnitude as $q_{\mathrm{Si}} = 1.7335\ |e|$ obtained in our all-electron DFT calculations with localized basis set; probably by chance the SQE/ESP predicted charge $q_{\mathrm{Si}} = 1.3824\ |e|$ nicely fits the data of ref.[70]

Magnitude of Hirshfeld-I charges. A general observation in the Hirshfeld-I results is that the absolute values of HI charges are significantly larger than the corresponding absolute values of the ESP fitted charges. This finding is markedly different from earlier studies where a reasonable correspondence between HI and ESP fitted charges was found for a set of organic molecules.[47] We suspect that this is a deficiency inherent to the Hirshfeld-I procedure when applied to (nearly) ionic systems. The oxygen charges in the oxide clusters are mostly between -1.0 and -2.0 $|e|$, which means that the corresponding proatoms in the Hirshfeld-I scheme are a linear interpolation between the isolated oxygen anion and dianion. The density profiles of these anionic pro-atoms are mostly determined by the limitations of the 6-311G+(d,p) basis set, which may lead to artifacts in

the Hirshfeld-I partitioning. In contrast, the charges on oxygen atoms in organic molecules are the range from 0 to -1.0 $|e|$ and therefore the basis set limitations are expected to be less important for constructing the pro-atoms.[26] This issue should be analyzed in detail in future work, potentially leading to an improved Hirshfeld-I scheme that also gives ESP-quality charges for oxides and ionic systems.

4 Conclusions

An extensive parametrization of the electronegativity equalization model (EEM) and split-charge equilibration (SQE) model was performed for silicate materials on the basis of quantum-chemical calculations of oxide clusters containing aluminum, silicon, and zirconium cations. The calibration of the parameters in these models was done using the iterative Hirshfeld (HI) charges and the electrostatic potential (ESP) as the reference quantities. The total cost function used in the non-linear least squares minimization procedure was a linear combination of a static cost function based on HI charges or ESP grid data, and a response cost function that included changes of these reference quantities upon an applied external electric field. The transferability of parameters was assessed by a comparison with HI charges and ESP grid data for a validation set of isolated molecules and a number of crystalline structures.

The outcome of the parametrization procedure allows us to conclude that the EEM model is capable of mimicking static characteristics only, while it fails to reproduce the response of the electronic distribution and ESP to an external electric field. The SQE model performs well for both static and response properties and also provides correct results for properties that were not explicitly included in the parameterization procedure, i.e. the dipole moment and the dipole polarizability.

Both the EEM and SQE calibrations can be used for a fast and reliable calculation of HI charges in isolated molecules that were not included in the training set, but they reveal a limited transferability to periodic systems. It is however possible to propose a correction to the atomic hardness

parameters of each element based on the differences between the EEM/HI and the DFT/HI charges for the periodic systems. This correction amounts to an increase of the atomic hardness in the solid state, which can be related to decrease of atomic polarizability due to the confinement by the crystal field. These corrections were found to be transferable between the EEM and SQE models. The corrected EEM/HI and SQE/HI schemes reproduce the reference DFT HI charges in the periodic systems with the mean relative error of less than 2 %.

The ESP-based parametrizations were found to provide reliable results only in those regions of space, where the electronic density values are lower than a threshold value $\rho = 10^{-3}$ a.u. The density criterion was found to be consistent with the distance criterion used in the Merz-Kollman ESP charge fitting scheme for a quick selection of grid points. The SQE/ESP model shows a good transferability from molecular to periodic structures, if the regions for computing ESP were chosen according to the criterion above. The EEM/ESP scheme was found to perform the worst among all models studied and hence the use of EEM method for computing the ESP-related characteristics should be avoided.

As an indirect result of this work, we observed very large differences between ESP fitted charges and iterative Hirshfeld charges. Future work should clarify the origin of the overestimation of the charge transfer in partially ionic systems by the iterative Hirshfeld scheme. Ideally, such work could lead to an improved iterative Hirshfeld scheme, whose charges satisfactorily approximate the electrostatic potential of silica clusters.

Acknowledgement

TV thanks the Fund for Scientific Research - Flanders (FWO). SS gratefully acknowledges the fellowship from the Ministère de l'Enseignement Supérieur et de la Recherche. The authors would also like to thank the Ghent University and the University of Lille 1 for allocation of the computational resources (Stevin Supercomputer Infrastructure and Centre de Ressources Informatiques, respectively). VVS acknowledges the European Research Council under the European Community Seventh Framework Programme (FP7(2007-2013) ERC grant agreement number 240483).

Supporting Information Available

XYZ files for the isolated systems in the training and validation set, and CIF files for the periodic systems. Figures with the λ-scans for the calibrations EEM/ESP and SQE/HI, and images of all isolated and periodic structures. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Mortier, W.; Ghosh, S.; Shankar, S. *J. Am. Chem. Soc.* **1986**, 108, 4315-4320.

(2) Rappe, A.; Goddard, W. *J. Phys. Chem.* **1991**, 95, 3358-3363.

(3) Chelli, R.; Procacci, P.; Righini, R.; Califano, S. *J. Chem. Phys.* **1999** 111, 8569.

(4) Nistor, R. A.; Polihronov, J. G.; Müser, M. H.; Mosey, N. J. *J. Chem. Phys.* **2006** 125, 094108.

(5) Yang, Z. Z.; Wang, C. S. *J. Phys. Chem. A* **1997**, 101, 6315-6321.

(6) Chen, J.; Martínez, T. J. *Chem. Phys. Lett*. **2007** 438, 315.

(7) York, D. M.; Yang, W. *J. Chem. Phys.* **1996**, 104, 159-172.

(8) Van Genechten, K. A.; Mortier, W. J.; Geerlings, P. *J. Chem. Phys.* **1987**, 86, 5063-5071.

(9) Janssens, G. O.; Baekelandt, B. G.; Toufar, H.; Mortier, W. J.; Schoonheydt, R. A. *J. Phys Chem.* **1995**, 99, 3251-3258.

(10) Baekelandt, B. G.; Mortier, W. J.; Lievens, J. L.; Schoonheydt, R. A. *J. Am. Chem. Soc.* **1991**, 113, 6730-6734.

(11) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, 101, 6141-6156.

(12) Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **2001**, 115, 2237-2251.

(13) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2004**, 120, 9903-9906.

(14) Chelli, R.; Schettino, V.; Procacci, P. *J. Chem. Phys.* **2004**, 122, 234107.

(15) Chelli, R.; Barducci, A.; Bellucci, L.; Schettino, V.; Procacci, P. *J. Chem. Phys.* **2004**, 123, 194109.

(16) Piquemal, J.-P.; Chelli, R.; Procacci, P.; Gresh, N. *J. Phys. Chem. A* **2007**, 111, 8170-8176.

(17) Tabacchi, G.; Mundy, C. J.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **2002**, 117, 1416-1433

(18) Chelli, R.; Ciabatti, S.; Cardini, G.; Righini, R.; Procacci, P. *J. Chem. Phys.* **1999**, 111, 4218-4229.

(19) Patel, S.; Brooks, C. L. *J. Comput. Chem.* **2004**, 25, 1-16.

(20) Chelli, R.; Pagliai, M.; Procacci, P.; Cardini, G.; Schettino, V. *J. Chem. Phys.* **2008**, 122, 074504.

(21) Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. B* **1999**, 103, 4730-4737.

(22) Chelli, R.; Procacci, P. *J. Chem. Phys.* **2002**, 117, 9175-9189.

(23) Bultinck, P.; Langenaeker, W.; Carbó-Dorca, R.; Tollenaere, J. P. *J. Chem. Inf. Comput. Sci.* **2003**, 43, 422-428.

(24) Masia, M.; Probst, M.; Rey, R. *J. Chem. Phys.* **2004**, 121, 7362-7378.

(25) Mathieu, D. The *J. Chem. Phys.* **2007**, 127, 224103

(26) Verstraelen, T.; van Speybroeck, V.; Waroquier, M. *J. Chem. Phys.* **2009**, 131, 044127.

(27) Mikulski, P. T.; Knippenberg, M. T.; Harrison, J. A. *J. Chem. Phys.* **2009**, 131, 241105.

(28) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, 110, 741-754.

(29) Bultinck, P.; Vanholme, R.; Popelier, P. L. A.; De Proft, F.; Geerlings, P. *J. Phys. Chem. A* **2004**, 108, 10359-10366.

(30) Smirnov, K. S.; van de Graaf, B. *J. Chem. Soc. Faraday Trans.* **1996**, 92, 2475-2480.

(31) Smirnov, K. S.; Thibault-Starzyk, F. *J. Phys. Chem. B* **1999**, 103, 8595-8601.

(32) Patel, S.; Mackerell, A. D.; Brooks, C. L. *J. Comput. Chem.* **2004**, 25, 1504-1514.

(33) Njo, S. L.; Fan, J. F.; van de Graaf, B. *J. Mol. Catal. A*, **1998**, 134, 79-88.

(34) Warren, G. L.; Davis, J. E.; Patel, S. *J. Chem. Phys.* **2008**, 128, 144110.

(35) Verstraelen, T.; Pauwels, E.; De Proft, F.; Van Speybroeck, V.; Geerlings, P.; Waroquier, M. Submitted to *J. Chem. Theory Comput.*

(36) Nistor, R. A.; Müser, M. H. *Phys. Rev. B* **2009**, 79, 104303.

(37) Weber, V.; Niklasson, A. M. N.; Challacombe, M. *Phys. Rev. Lett.* **2004**, 92, 193002.

(38) Krishtal, A.; Senet, P.; Yang, M.; Van Alsenoy, C. *J. Chem. Phys.* **2006**, 125, 034312.

(39) Kirtman, B.; Toto, J. L.; Robins, K. A.; Hasan, M. *J. Chem. Phys.* **1995**, 102, 5350-5356.

(40) van Faassen, M.; de Boeij, P. L.; van Leeuwen, R.; Berger, J. A.; Snijders, J. G. *Phys. Rev. Lett.* **2002**, 88, 186401.

(41) Parr, R. G.; Ayers, P. W.; Nalewajski, R. F. *J. Phys. Chem. A* **2005**, 109, 3957-3959.

(42) Jiroušková, Z.; Svobodová Vařeková, R.; Vaněk, J.; Koča, J. *J. Comput. Chem.* **2009**, 30, 1174-1178.

(43) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Van Alsenoy, C.; Tollenaere, J. P. *J. Phys. Chem. A* **2002**, 106, 7895-7901.

(44) Geerlings, P.; De Proft, F.; Langenaeker, W. *J. Chem. Rev.* **2003**, 103, 1793-1873.

(45) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Dorca, R. C. *J. Chem. Phys.* **2007**, 126, 144111.

(46) Bultinck, P.; Ayers, P. W.; Fias, S.; Tiels, K.; Van Alsenoy, C. *Chem. Phys. Lett.* **2007**, 444, 205-208.

(47) Van Damme, S.; Bultinck, P.; Fias, S. *J. Chem. Theory Comput.* **2009**, 5, 334-340.

(48) Catak, S.; D'hooghe, M.; Verstraelen, T.; Hemelsoet, K.; Van Nieuwenhove, A.; Ha, H.-J.; Waroquier, M.; De Kimpe, N.; Van Speybroeck, V. *J. Org. Chem.* **2010**, 75, 4530-4541.

(49) Baerlocher, C.; McCusker, L.; Olson, D. *Atlas of Zeolite Framework Types*, 6th ed.; Elsevier: Amsterdam, The Netherlands, **2007**; pp 1-398. http://www.iza-structure.org/databases/

(50) Becke, A. D. *J. Chem. Phys.* **1993**, 98, 5648-5652.

(51) Raghavachari, K.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, 72, 650-654.

(52) McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, 72, 5639-5648

(53) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, 82, 270-283.

(54) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.;

Wong, M. W.; Gonzalez, C.; and Pople, J. A. *Gaussian 03, Revision D.01*, Gaussian, Inc., Wallingford CT, 2004..

(55) Dovesi, R.; Saunders, V. R.; Roetti, C.; Orlando, R.; Zicovich-Wilson, C. M.; Pascale, F.; Civalleri, B.; Doll, K.; Harrison, N. M.; Bush, I. J.; D'Arco, P.; Llunell, M. *CRYSTAL06 User's Manual*. University of Torino: Torino, 2006.

(56) Gatti, C.; Saunders, V.R.; Roetti, C.; *J. Chem. Phys.* **1994**, 101, 10686-10696.

(57) Catti, M.; Valerio, G.; Dovesi R.; M. Causà, M. *Phys. Rev. B* **1994**, 49, 14179-14187.

(58) Pascale, F.; Zicovich-Wilson, C. M.; Orlando, R.; Roetti, C.; Ugliengo, P.; Dovesi, R. *J. Phys. Chem. B* **2005**, 109, 6146-6152.

(59) http://www.tcm.phy.cam.ac.uk/~mdt26/crystal.html

(60) Monkhorst, H. J.; Pack, J. D. *Phys. Rev. B* **1976** 13, 5188.

(61) http://molmod.ugent.be/code/wiki/HiPart

(62) Becke, A. D. *J. Chem. Phys.* **1988**, 88, 2547-2553.

(63) Verstraelen, T.; Bultinck, P.; Van Speybroeck, V.; Ayers, P. W.; Van Neck, D.; Waroquier, M. *J. Chem. Theory Comput.* **2011**, 7, 1750-1764.

(64) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. Conjugate Gradient Methods in Multidimensions. In *Numerical Recipes in C: The Art of Scientific Computing*; Cambridge University Press: New York, **1992**; pp 420-424.

(65) Politzer, P. *J. Chem. Phys.* **1987**, 86, 1072.

(66) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, 5, 129-145.

(67) Mantina, M.; Chamberlin, A. C.; Valero, R.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2009**, 113, 5806-5812.

(68) Kokalj, A. *Comp. Mater. Sci.* **2003**, 28, 155. Code available from http://www.xcrysden.org/.

(69) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, 97, 10269-10280.

(70) Campanã, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, 5, 2866.

(71) Ewald, P. *Ann. Phys.* **1921**, 369, 253-287.

(72) Smirnov, K. S.; Bougeard, D. *Chem. Phys.* **2003**, 292, 53.

Appendices

A Hardness matrix elements of the EEM and SQE models for periodic systems

For periodic systems, the electrostatic interactions are generally treated with the Ewald sumation.[71] A combination of the method with a screened Coulombic potential, see for example Eq. (7), results in the following expression for the electrostatic interaction energy:[72]

$$\begin{aligned} E_{el} = &\frac{1}{2}\sum_i \sum_{j\neq i} q_i q_j \Big[J(r_{ij}) - \frac{\mathrm{erf}(\beta r_{ij})}{r_{ij}} \Big] \\ &+ \frac{2\pi}{V} \sum_{\mathbf{k}\neq\mathbf{0}} \frac{e^{-|\mathbf{k}|^2/4\beta^2}}{|\mathbf{k}|^2} \sum_i \sum_j q_i q_j e^{i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} - \frac{\beta}{\sqrt{\pi}} \sum_i q_i^2, \end{aligned} \tag{A.1}$$

where $J(r)$ is the screened electrostatic potential, and β controls the convergence of the Ewald sums in the direct and reciprocal space. For periodic systems Eq. (A.1) replaces the last term in Eq. (3) and Eq. (6).

Making use of Eq. (3), Eq. (6), and Eq. (A.1) the charge-dependent energy of periodic system in the EEM and SQE models can be rewritten as

$$E_{\mathrm{EEM}}(\mathbf{q}) = \sum_i \big[\chi_i q_i + \frac{1}{2}\tilde{\eta}_i q_i^2\big] + \frac{1}{2}\sum_i \sum_{j\neq i} q_i q_j J'(r_{ji}), \tag{A.2}$$

$$\begin{aligned} E_{\mathrm{SQE}}(\mathbf{p}) = &\sum_i \sum_k^{(i)} \xi_{ik} p_{ik} + \frac{1}{2}\sum_i \sum_k^{(i)} \tilde{\kappa}_{ik} p_{ik}^2 + \frac{1}{2}\sum_i \tilde{\eta}_i \sum_k^{(i)} \sum_{l\neq k}^{(i)} p_{ik} p_{il} \\ &+ \frac{1}{2}\sum_i \sum_{j\neq i} J'(r_{ij}) \sum_k^{(i)} \sum_l^{(j)} p_{ik} p_{jl}, \end{aligned} \tag{A.3}$$

where the superscript (i) is used to denote that the sum is limited to those atoms that are "allowed"

to exchange charge with atom (i). The quantities $\tilde{\eta}_i$, $\tilde{\kappa}_i$ and $J'(r_{ji})$ stand for

$$\tilde{\eta}_i = \eta_i + \frac{4\pi}{V}\sum_{\mathbf{k}\neq\mathbf{0}} \frac{e^{-|\mathbf{k}|^2/4\beta^2}}{|\mathbf{k}|^2} - \frac{2\beta}{\sqrt{\pi}} \tag{A.4}$$

$$\tilde{\kappa}_{ji} = \kappa_{ji} + \tilde{\eta}_i \tag{A.5}$$

$$J'(r_{ji}) = J(r_{ij}) - \frac{\operatorname{erf}(\beta r_{ij})}{r_{ij}} + \frac{4\pi}{V}\sum_{\mathbf{k}\neq\mathbf{0}} \frac{e^{-|\mathbf{k}|^2/4\beta^2}}{|\mathbf{k}|^2} e^{i\mathbf{k}\cdot(\mathbf{r}_i-\mathbf{r}_j)} \tag{A.6}$$

As before, a generic formula for hardness matrix element $H_{(ji)(lm)}$ in SQE model can be obtained by the differentiation of Eq. (A.3) with respect to the CTPs p_{ji} and p_{lm}

$$\begin{aligned} H_{(ji)(lm)} = {} & \delta_{jl}\delta_{im}\left(\tilde{\kappa}_{ji} + \tilde{\kappa}_{ij} - 2J'_{ji}\right) + \left[\delta_{im}(1-\delta_{jl})\tilde{\eta}_i - \delta_{il}\tilde{\eta}_i - \delta_{jm}\tilde{\eta}_j + \delta_{jl}(1-\delta_{im})\tilde{\eta}_j\right] \\ & + \left(1-\delta_{jl}\delta_{im}\right)\left[\left(1-\delta_{jl}\right)J'_{jl} - \left(1-\delta_{il}\right)J'_{il} - \left(1-\delta_{jm}\right)J'_{jm} + \left(1-\delta_{im}\right)J'_{im}\right], \end{aligned} \tag{A.7}$$

where $J'_{ji} \equiv J'(r_{ji})$ and δ_{ji} is the Kronecker symbol.

B Static and response cost functions

External perturbation. In addition to the static quantities (A), such as HI charges and electrostatic potential, a their response to an external perturbation V_{ext} was computed and the resulting response characteristics were then employed in constructing the corresponding response cost function. The external perturbation was given by

$$V_{\text{ext}} = \mathscr{E}_x x + \mathscr{E}_y y + \mathscr{E}_z z \tag{B.1}$$

where $\mathscr{E}_\alpha$ $(\alpha = x,y,z)$ is an applied uniform electric field in x,y and z directions. The derivative of quantity A with respect to the perturbation $\mathscr{E}_\alpha$ was obtained by the finite differences as

$$A'_\alpha \approx \frac{A(\mathscr{E}_\alpha = +\varepsilon) - A(\mathscr{E}_\alpha = -\varepsilon)}{2\varepsilon}. \tag{B.2}$$

A value of $\varepsilon = 0.0019$ a.u. is used, which is equal to the default value in Gaussian03 code when a numerical differentiation is carried out with respect to external electric field.

Static HI cost function. The static HI cost function uses the iterative Hirshfeld charges and it has the following form

$$X_{\mathrm{S,HI}} = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N_m} w_{mn}\left[q_{mn}^{0} - q_{mn}^{\mathrm{MODEL}}\right]^2}{\sum_{m=1}^{M}\sum_{n=1}^{N_m} w_{mn}}, \tag{B.3}$$

where the sums in the numerator run over M molecules and N_m atoms in a given molecule, respectively; the superscript "0" denotes the reference value and MODEL stands for value obtained with either EEM or SQE model. The weight w_{mn} associated with a certain atomic charge is inverse proportional to the prevalence of the corresponding chemical element in the training set (Table 1). This choice compensates for the fact that some chemical elements are more abundant than others. The cost function (Eq. (B.3)) can be seen as a weighted average of the individual squared errors.

Response HI cost function. The response cost function related to the HI charges takes the following form:

$$X_{\mathrm{R,HI}} = \frac{\sum_{m=1}^{M}\sum_{\alpha=x,y,z}\sum_{n=1}^{N_m} w_{mn}\left[\left(q_{mn}^{0}\right)'_{\alpha} - \left(q_{mn}^{\mathrm{MODEL}}\right)'_{\alpha}\right]^2}{\sum_{m=1}^{M}\sum_{n=1}^{N_m} 3w_{mn}} \tag{B.4}$$

The derivatives of the model HI charges towards the external field are computed analytically. The weights are identical to those of the static HI cost function (Eq. (B.3)).

Static ESP cost function. For every molecule m, a set G_m of grid points $\mathbf{r}_{m,g}$ ($g \in G_m$) with their associated weights $w_{m,g}$ was defined. At these grid points the DFT and model electrostatic

potentials are computed and used to construct static ESP cost function as follows:

$$X_{\mathrm{S,ESP}} = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{\sum_{g=1}^{G_m} w_{m,g}\left[V_m^0(\mathbf{r}_{m,g}) - V_m^{\mathrm{MODEL}}(\mathbf{r}_{m,g})\right]^2}{\sum_{g=1}^{G_m} w_{m,g}}\right). \tag{B.5}$$

The grid consists of an uniform 3D grid whose edges are separated by at least 5 Å from the nearest atom. The spacing between the grid points is 0.2 Å. The part between brackets in eq (Eq. (B.5)) represents an ESP cost function for a molecule m. The evaluation of this molecular cost function is made more efficient by reducing it to a simple quadratic function of charges prior to the calibration procedure.

The weight function $w_m(\mathbf{r}_{m,g})$ is smooth, becomes one in a shell surrounding the molecule and decays to zero inside the molecule and at large distances, i.e. at the edges of the grid. This weight function is constructed as the product of two switching functions:

$$w(\mathbf{r}) = \mathrm{switch}\left[-\alpha(\log_{10}(\rho(\mathbf{r})) - \log_{10}(\rho_0))\right] \times \mathrm{switch}\left[-\beta(\mathrm{d}(\mathbf{r}) - d_0)\right] \tag{B.6}$$

where $\mathrm{switch}(x)$ and $\mathrm{d}(\mathbf{r})$ are defined as

$$\mathrm{switch}(x) = \begin{cases} 0 & \text{if } x < -1 \\ (1+\sin(\pi x/2))/2 & \text{if } -1 < x < 1 \\ 1 & \text{if } x > 1 \end{cases} \tag{B.7}$$

$$\mathrm{d}(\mathbf{r}) = -a_0 \log\left[\sum_{j=1}^{N} \exp(-\|\mathbf{R}_j - \mathbf{r}\|/a_0)\right]. \tag{B.8}$$

The first factor in the weight function switches to zero inside the molecule. The second factor switches of the weight function at long distances. The function $d(\mathbf{r})$ is a smooth function that approximates the distance to the closest atom. The parameters are fixed to α = 1.0, $\rho_0 = 3 \cdot 10^{-4}$ a.u., $\beta = 0.5$ Å, $d_0 = 4.0$ Å, and a_0 is the Bohr radius.

The smoothness of the weight function guarantees that a small rotation of the molecules with respect to the grid only leads to small changes in the cost function. A second advantage of this approach is that it never includes grid points in the fitting procedure where the electron density is significantly non-zero.

Response ESP cost function. The response ESP cost function takes the following form:

$$X_{\mathrm{R,ESP}} = \frac{1}{3M} \sum_{m=1}^{M} \sum_{\alpha=x,y,z} \left(\frac{\sum_{g=1}^{G_m} w_{m,g} \left[\left(V_m^0\right)'_\alpha(\mathbf{r}_{m,g}) - \left(V_m^{\mathrm{MODEL}}\right)'_\alpha(\mathbf{r}_{m,g}) \right]^2}{\sum_{g=1}^{G_m} w_{m,g}} \right). \tag{B.9}$$

The derivatives of the model ESP towards the external field was computed numerically. The weights were identical to those in the static ESP cost function (Eq. (B.5)). Similar to the static ESP cost function, the part between brackets was first reduced to a quadratic function of charges for each molecule to lower the computational cost of the calibration procedure.

Table 1: Number of atoms of each element in training and validation sets for molecular systems.

Atom	Training	Validation
H	747	747
O	534	533
Al	57	57
Zr	65	65
Si	230	231
Total	1633	1633

Table 2: Structural characteristics of periodic systems.

Structure	Space group	No. of atoms per unit cell	V_0 (Å^3)[a]	FD[b]
α-quartz	$P3_221$	9	113.007	26.6
α-cristobalite	$P4_12_12$	12	171.104	23.4
JBW (SiO_2)	*Pmma*	18	321.279	18.7
DFT (SiO_2)	$P4_2/mmc$	24	456.609	17.5
NPO (SiO_2)	$P\bar{6}2c$	18	350.960	17.1
SOD (SiO_2)	*Im*3*m*	36	725.049	16.6
t-ZrO_2	$P4_2/nmc$	6	67.810	
$ZrSiO_4$ (zircon)	$I4_1/amd$	24	269.620	

[a] Unit cell volume after geometry optimization. [b] Framework density (number of Si atoms per 1000 Å^3).

Table 3: EEM and SQE parameters obtained in the NLLSQ fits.

Parameter	EEM-HI	EEM-ESP	SQE-HI	SQE-ESP
χ_{H} (eV)	0.0000	0.0000		
χ_{O} (eV)	4.1441	3.7764		
χ_{Al} (eV)	−7.0722	−5.9772		
χ_{Si} (eV)	−5.4968	−0.1682		
χ_{Zr} (eV)	−17.3596	−221.4566		
η_{H} (eV)	13.7726	16.7508	11.9294	14.4867
η_{O} (eV)	15.0954	17.5301	11.9075	10.6047
η_{Al} (eV)	12.1315	15.0322	10.4054	10.0619
η_{Si} (eV)	11.8331	9.4161	10.2542	9.2006
η_{Zr} (eV)	13.7565	153.5238	10.7434	8.5684
$\xi_{\mathrm{H\text{-}O}}$ (eV)			−2.7666	−1.9923
$\xi_{\mathrm{H\text{-}Al}}$ (eV)			3.1308	2.3323
$\xi_{\mathrm{H\text{-}Si}}$ (eV)			2.0820	0.4558
$\xi_{\mathrm{H\text{-}Zr}}$ (eV)			5.9027	1.6812
$\xi_{\mathrm{O\text{-}Al}}$ (eV)			7.7564	4.6836
$\xi_{\mathrm{O\text{-}Si}}$ (eV)			5.8791	1.6479
$\xi_{\mathrm{O\text{-}Zr}}$ (eV)			8.3020	2.3925
$\kappa_{\mathrm{H\text{-}O}}$ (eV)			5.1124	3.1167
$\kappa_{\mathrm{H\text{-}Al}}$ (eV)			3.1262	1.0450
$\kappa_{\mathrm{H\text{-}Si}}$ (eV)			3.7648	1.3472
$\kappa_{\mathrm{H\text{-}Zr}}$ (eV)			2.5118	0.1420
$\kappa_{\mathrm{O\text{-}Al}}$ (eV)			6.1704	5.8499
$\kappa_{\mathrm{O\text{-}Si}}$ (eV)			6.1868	4.1075
$\kappa_{\mathrm{O\text{-}Zr}}$ (eV)			4.0558	2.9940
R_{H} (Å)	0.7724	0.4850	0.6810	0.5608
R_{O} (Å)	0.5382	0.4634	0.6823	0.7661
R_{Al} (Å)	0.7230	0.5405	0.7833	1.0523
R_{Si} (Å)	0.8082	0.8913	0.7923	0.8924
R_{Zr} (Å)	0.8551	0.0529	0.9414	1.0305

Table 4: Atomic charges in periodic systems obtained with different schemes.

Structure	EEM/HI	SQE/HI	DFT/HI	EEM/ESP	SQE/ESP	DFT/ESP
			O atoms			
α-quartz	−1.6417	−1.5750	−1.2840	−0.9800	−0.6833	−0.7741
α-cristobalite	−1.6281	−1.5715	−1.2762	−0.9634	−0.6839	−1.0512
JBW	−1.6605	−1.5902	−1.2933	−0.9681	−0.6924	−0.8167
	−1.6894	−1.6122	−1.2974	−0.9801	−0.7047	−0.7978
	−1.6114	−1.5620	−1.2846	−0.9588	−0.6742	−0.7622
	−1.6337	−1.5825	−1.3001	−0.9811	−0.6858	−0.8705
DFT	−1.6273	−1.5708	−1.2805	−0.9581	−0.6814	−0.8818
	−1.6941	−1.6184	−1.3081	−0.9877	−0.7093	−0.9121
	−1.5624	−1.5245	−1.2615	−0.9284	−0.6531	−0.7805
SOD	−1.6526	−1.5912	−1.2966	−0.9733	−0.6911	−0.8667
NPO	−1.6060	−1.5549	−1.2594	−0.9401	−0.6759	−0.6574
	−1.5641	−1.5239	−1.2591	−0.9204	−0.6549	−0.5393
t-ZrO2	−1.7224	−1.7095	−1.5542	−0.7721	−0.9574	−1.7357
ZrSiO4	−1.6363	−1.8660	−1.3914	−0.8453	−1.0227	−1.2971
			Si atoms			
α-quartz	3.2834	3.1500	2.5680	1.9600	1.3667	1.5482
α-cristobalite	3.2562	3.1430	2.5524	1.9268	1.3678	2.1024
JBW	3.2648	3.1522	2.5707	1.9020	1.3684	1.5716
	3.3373	3.1945	2.6118	2.0110	1.3869	1.6829
DFT	3.2555	3.1422	2.5654	1.9161	1.3626	1.7281
SOD	3.3052	3.1824	2.5932	1.9466	1.3822	1.7335
NPO	3.1701	3.0788	2.5185	1.8605	1.3308	1.1967
ZrSiO4	3.0416	3.1342	2.4133	1.8404	1.5285	2.1799
			Zr atoms			
t-ZrO2	3.4447	3.4190	3.1084	1.5442	1.9148	3.4714
ZrSiO4	3.5037	4.3298	3.1525	1.5409	2.5622	3.0086

Table 5: Elements of G matrix (cf. Eq. (17)) computed for the oxygen and silicon atoms in crystalline silica structures using the parameters of EEM/HI scheme.

Structure	G_{O} (eV)	G_{Si} (eV)
α-quartz	0.8827	0.3761
α-cristobalite	0.8196	0.4215
NPO	0.6264	0.4729
SOD	0.7388	0.4313

Table 6: Hirshfeld-I atomic charges computed with the molecular EEM/HI and SQE/HI models, and with the models corrected for solid state calculations (EEM_S/HI and SQE_S/HI); the right-hand column reports the reference Hirshfeld-I charges obtained in periodic DFT computations (DFT/HI).

Structure	EEM/HI	EEM_S/HI	SQE/HI	SQE_S/HI	DFT/HI
			O atoms		
α-quartz	−1.6417	−1.2739	−1.5750	−1.2956	−1.2840
α-cristobalite	−1.6281	−1.2881	−1.5715	−1.2932	−1.2762
JBW	−1.6605	−1.2869	−1.5902	−1.3068	−1.2933
	−1.6894	−1.3081	−1.6122	−1.3239	−1.2974
	−1.6114	−1.2518	−1.5620	−1.2846	−1.2846
	−1.6337	−1.2663	−1.5825	−1.2998	−1.3001
DFT	−1.6273	−1.2644	−1.5708	−1.2927	−1.2805
	−1.6941	−1.3141	−1.6184	−1.3293	−1.3081
	−1.5624	−1.2185	−1.5245	−1.2570	−1.2615
SOD	−1.6526	−1.3153	−1.5912	−1.3065	−1.2966
NPO	−1.6060	−1.2544	−1.5549	−1.2833	−1.2594
	−1.5641	−1.2263	−1.5239	−1.2595	−1.2591
			Si atoms		
α-quartz	3.2834	2.5478	3.1500	2.5912	2.5680
α-cristobalite	3.2562	2.5762	3.1430	2.5864	2.5524
JBW	3.2648	2.5319	3.1522	2.5908	2.5707
	3.3373	2.5880	3.1945	2.6249	2.6118
DFT	3.2555	2.5307	3.1422	2.5858	2.5654
SOD	3.3052	2.6306	3.1824	2.6131	2.5932
NPO	3.1701	2.4807	3.0788	2.5428	2.5185

Graphical TOC Entry